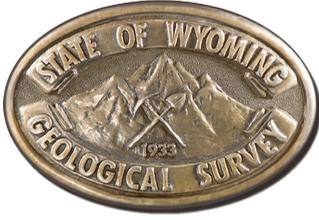


Interpreting the past, providing for the future

Regional-Scale Geochemical Investigations from Legacy Rock and Sediment Datasets

By Jesse R. Pisel and Charles P. Samra

Open File Report 2019-2
April 2019



Wyoming State Geological Survey

Erin A. Campbell, Director and State Geologist



Regional-Scale Geochemical Investigations from Legacy Rock and Sediment Datasets

By Jesse R. Pisel and Charles P. Samra

Layout by Christina D. George

Open File Report 2019-2
Wyoming State Geological Survey
Laramie, Wyoming: 2019

This Wyoming State Geological Survey (WSGS) Open File Report is preliminary and may require additional compilation and analysis. Additional data and review may be provided in subsequent years. For more information about the WSGS, or to download a copy of this Open File Report, visit www.wsgs.wyo.gov. The WSGS welcomes any comments and suggestions on this research. Please contact the WSGS at 307-766-2286, or email wsgs-info@wyo.gov.

Citation: Pisel, J.R., and Samra, C.P., 2019, Regional-scale geochemical investigations from legacy rock and sediment datasets: Wyoming State Geological Survey Open File Report 2019-2, 20 p.

Table of Contents

Abstract	1
Introduction	1
Background	1
Sample Types	1
National Uranium Resource Evaluation Sediment Data	2
National Geochemical Database Rock Data	2
Hydrologic Units	2
Methods	5
Multiple Imputation by Chained Equations	5
Spatial Associations	9
Getis-Ord G_i^* Hotspot Analysis	10
G_i^* Values	10
Classification	10
Classification Validation Workflow	12
Results	13
Titanium	13
Rare Earth Elements	13
Validation	13
Case Studies	16
Titanium	16
Rare Earth Elements	18
Future Sampling	18
Summary and Conclusions	18
References	20

List of Figures

Figure 1. Map showing locations of NURE sediment samples and NGDBR rock-chip samples.	3
Figure 2. Data-dense matrices that document patterns in missing data.	4
Figure 3. Map of the hydrologic unit subwatershed basins across Wyoming.	5
Figure 4. Comparison of imputed and actual values for lutetium, niobium, and titanium.	8
Figure 5. Maps of contiguity spatial networks used in hotspot analysis.	9
Figure 6. Visual representation of the classification scheme	11
Figure 7. Map showing results from hotspot analysis and classification for titanium	14

Figure 8. Map showing results from hotspot analysis and classification for all rare earth elements15

Figure 9. Map of results from case studies17

Figure 10. Map of case-study results from the Greater Green River Basin19

ABSTRACT

Understanding the regional distribution of elements is important for optimizing mineral sampling programs at large and small scales. This study takes advantage of sediment samples from the National Uranium Resource Evaluation program and rock-chip samples from the National Geochemical Database to predict regional changes in relative abundance of elements in Wyoming. We use multiple imputation by chained equations to assign values to the incomplete geochemical data suite of these two historical programs, resulting in a complete geochemical dataset of 48 elements. From the complete dataset, we use hotspot analysis to predict areas of interest for both the sediment and rock data; areas are defined by watersheds. We then classify each watershed in the state by mineralization potential. The classification model is validated using rare earth elements and titanium as case studies using independent samples not included in the model. We find the model tends to overestimate the number of unsampled mineralized areas, which we interpret as more beneficial than underestimation. The model does well in predicting areas of known and potential mineralization, and provides a way to rapidly prioritize areas of interest and streamline sampling programs at the regional scale.

INTRODUCTION

Geologic knowledge spanning an entire state expedites finding and exploiting new mineral resources. Implementing statewide rock-sampling programs to achieve this knowledge can be costly. However, geochemical data covering a substantial portion of the state are freely available from historical sampling programs. These geochemical data provide insight into areas of interest for potential resource development without requiring an in-depth geologic understanding of the area. We use spatial statistics and a classification scheme to document areas with known mineralization, missed mineralization, and potential mineralization throughout Wyoming. This classification highlights specific areas for focused sampling programs and further in-depth geologic mapping, thereby reducing time and cost to evaluate mineral resources across the state. The results of classification and hotspot analysis for all 48 elements are available online through the Wyoming State Geological Survey's [Mines and Minerals Map](http://wsgs.maps.arcgis.com/apps/webappviewer/index.html?id=af948a51f4954a81adeae8935440cd28) (<http://wsgs.maps.arcgis.com/apps/webappviewer/index.html?id=af948a51f4954a81adeae8935440cd28>) and [publication search page](https://www.wsgs.wyo.gov/pubs-maps/publication-search) (<https://www.wsgs.wyo.gov/pubs-maps/publication-search>).

BACKGROUND

Sample Types

When evaluating and interpreting geochemical data, it is important to understand what each sample type represents, recognizing the differences as well as the limitations and benefits of each type. Two sample types were used in this study: stream-sediment samples collected as part of the National Uranium Resource Evaluation program and rock samples from the National Geochemical Database. Stream-sediment samples are an amalgamation of detritus sourced from the bedrock within a given drainage basin and transported down gradient through geomorphic processes (alluvial, fluvial, eolian, etc.). Stream-sediment samples provide a first-order vector for the source of potential mineralization. The value of elemental concentrations should produce both positive and negative trends throughout a drainage basin with distance from the potential source of an anomaly due to the mixing of the sediments by surface processes. This sample type is minimally biased by the human sampler and represents a more heterogeneous sample over a broad region. Rock samples represent elemental values for a specific point location at the outcrop scale, and may give an actual representative value of mineralization in the region or even at the outcrop scale. Rock samples are extremely susceptible to sampler bias and heterogeneity, and may be subject to both “high grading” and “low grading.” Therefore, it is imperative to integrate the data of these two sample types in concert when evaluating a potentially mineralized area.

National Uranium Resource Evaluation Sediment Data

The National Uranium Resource Evaluation (NURE) program began in 1973 with a directive from the U.S. Atomic Energy Commission to identify uranium resources throughout the United States. A significant part of the program was directed at sampling, including stream sediment, soil, surface water, and groundwater. Sampling was conducted in most states, including Wyoming. Although the original goal of the project was to identify uranium resources, the samples were also analyzed for an additional 47 elements (Smith, 1997). These historic data are available from the U.S. Geological Survey (Smith, 1997).

In this study, we used data from 18,424 stream-sediment samples collected from March 1976 through October 1979 throughout Wyoming as part of the NURE program. The sediment samples span the entire state, with only three areas of sparse data coverage centered on the Bighorn Mountains in the north-central portion of the state, east of Cheyenne in the southeast, and around Riverton in central Wyoming (fig. 1). Sediments were collected using a clean polyethylene scoop and placed in a new polyethylene bag (Puchlik, 1977). After collection, the sediments were dried in ovens at temperatures of 100°C or less and then sieved using either an 80 or 100 mesh sieve, yielding sediments in the range of 149 to 177 micrometers (μm) in diameter for geochemical analysis (Sharp Jr. and Aamodt, 1978).

The sediment samples were analyzed between 1976 and 1980 for elemental composition using different methods at Los Alamos and Oak Ridge National laboratories. The analytical methods conducted at Los Alamos National Laboratory were delayed-neutron counting analysis, dispersive X-ray fluorescence (XRF), arc-source emission spectrography, and neutron activation. The analytical methods conducted at Oak Ridge National Laboratory were fluorescence spectroscopy, neutron activation, emission spectrochemical, atomic absorption spectroscopy, mass spectrometry, gamma spectrometer, atomic absorption spectroscopy for tin, and atomic absorption spectroscopy for mercury. There are samples within the dataset without results for the full suite of elements due to the inconsistent methods of analysis (fig. 2a).

National Geochemical Database Rock Data

The National Geochemical Database: Rock (NGDBR) data is a compilation of rock geochemistry data collected by the USGS since the early 1960s (U.S. Geological Survey, 2018). The rocks in this dataset were collected and analyzed as early as April 1964 and as recently as September 2001. The rock samples were collected from a variety of sources such as outcrops (81.9 percent), drill cores (7.8 percent), prospect pits (3.2 percent), mines and quarries (3.0 percent), and other unidentified sources (4.1 percent). This dataset contains 20,432 samples from Wyoming and spans most of the state (fig. 1). NGDBR rock sampling techniques are generally described as the rock being sampled from its source using a clean rock hammer and placed in a canvas or polyethylene bag until analysis (U.S. Geological Survey, 2018).

Analyses of the 20,432 samples are split into seven different methods: XRF spectrometry, neutron activation, inductively coupled plasma-atomic emission spectrometry, inductively coupled-mass spectrometry, emission spectroscopy, and other and unknown analytical methods. Similar to the NURE sediment geochemistry data, the large variety of analytical methods resulted in samples without analyses for the full suite of elements (fig 2b). Both the sediment and rock sample datasets have spatial location data for each sample that can be used to document spatial changes in elemental concentrations.

Hydrologic Units

Because the NURE dataset is composed of sediment samples, we chose to investigate spatial changes of element concentration by associating sample locations with drainage basins. We used drainage basins as the spatial framework because the sediment is transported through these drainage basins; we expect to observe changes in concentrations throughout drainage basins as elemental concentrations change with distance from potential sources. Both rock and sediment samples can be spatially associated with unique drainage basins, or hydrologic units, across the state. The U.S. Department of Agriculture created a watershed boundary dataset of hydrologic units defined as the areal

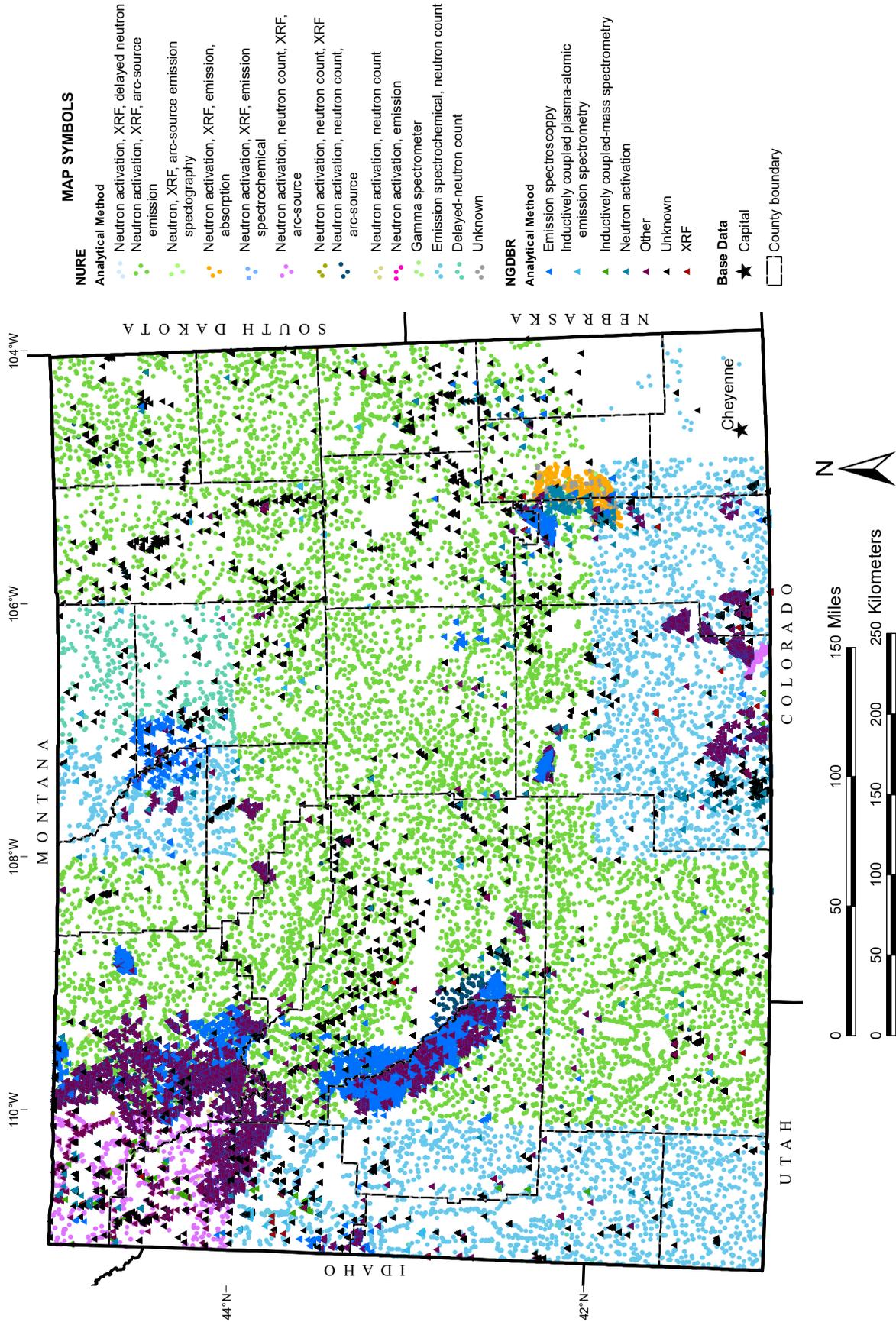


Figure 1. Map showing locations of NURE sediment samples and NGDBR rock-chip samples. Each sample location is symbolized by the analytical methods.

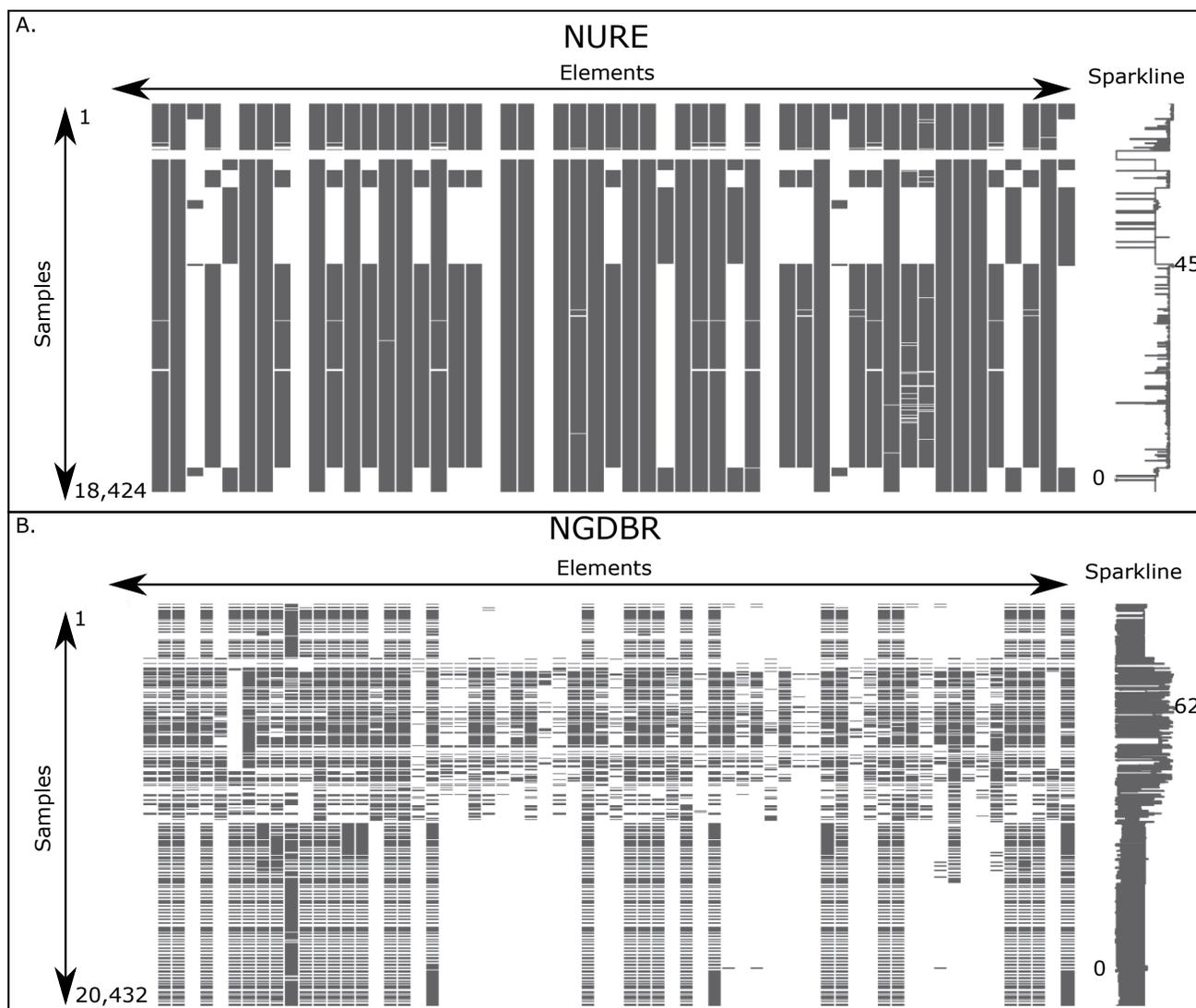


Figure 2. Data-dense matrices that document patterns in missing data. Columns correspond to the elements in the geochemical datasets, and rows correspond to each sample in the dataset. The sparkline on the right summarizes the general shape of the data completeness, and points out the maximum and minimum rows. When the sparkline shifts right it signifies more elements are present for that particular sample. The NURE dataset (A) contains fewer samples and fewer elements than the NGDBR dataset (B), but is more complete.

extent of surface water draining to a certain point. Furthermore, the Federal Standard for Delineation of Hydrologic Unit Boundaries defines a hydrologic unit (HU) as:

“a drainage area delineated to nest in a multi-level, hierarchical drainage system. Its boundaries are defined by hydrographic and topographic criteria that delineate an area of land upstream from a specific point on a river, stream, or similar surface waters. A hydrologic unit can accept surface water directly from upstream drainage areas, and indirectly from associated surface areas such as remnant, non-contributing, and diversions to form a drainage area with single or multiple outlet points. Hydrologic units are only synonymous with classic watersheds when their boundaries include all the source area contributing surface water to a single defined outlet point”
 (U.S. Geological Survey and U.S. Department of Agriculture, 2013).

HUs are given unique identifiers called a hydrologic unit code (HUC) that consist of a combination of region, sub-region, basin, subbasin, watershed, and subwatershed, depending on the hierarchical level of the hydrologic unit. In

Wyoming, there are 4 regional HUs, 14 subregional HUs, 17 basin HUs, 82 subbasin HUs, 422 watershed HUs, and 2,382 subwatershed HUs. In this study, we focus on the highest spatial resolution—the subwatershed—to limit future sampling programs to the smallest possible area. The 2,382 subwatersheds in Wyoming (fig. 3) have a mean area of 28,239 acres (114.2 km²). Yellowstone Lake, within the Yellowstone Caldera in northwestern Wyoming, is the largest subwatershed at 210,613 acres (852 km²). The smallest subwatershed is an unnamed depression near Fontenelle Reservoir that is 3,056 acres (12 km²) and is located in the Green River Basin in west-central Wyoming. Subwatersheds are the critical spatial framework for assessing the areal distribution of elements across the state.

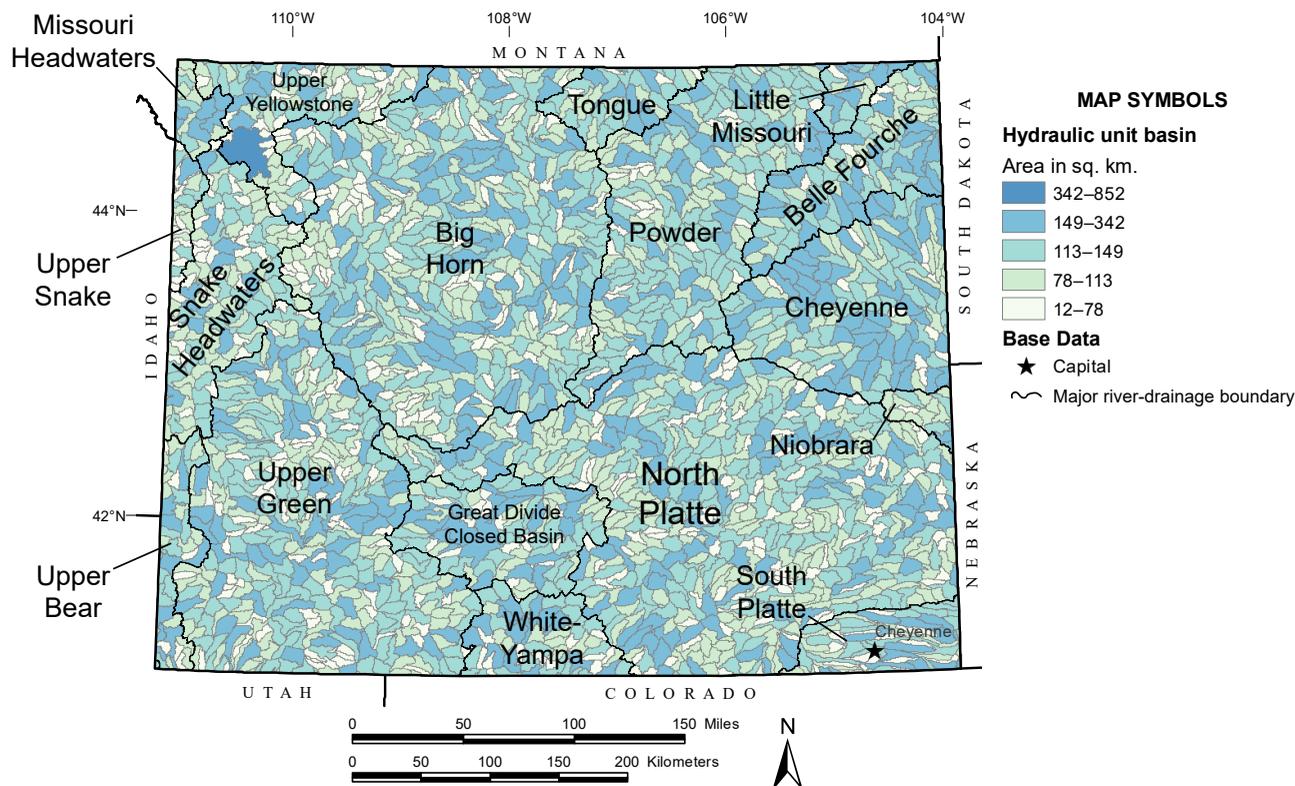


Figure 3. Map of the hydrologic unit subwatershed basins across Wyoming. The larger subregional-level hydrologic units are labeled and delineated by the dark black line.

METHODS

Multiple Imputation by Chained Equations

To populate the subwatersheds with elemental data, we first addressed the geochemical sample incompleteness. Because of the variety of analytical methods used to evaluate both the sediment and rock samples, the completeness of geochemical data for each sample is variable. Elemental analyses for some samples returned results below the instrument's detection limits, and the results for that element were recorded as the minimum detection limit value. Furthermore, all elements were not analyzed for all samples.

To address the issue of detection limits, we first reviewed results for individual samples with elements below the detection limit. For samples below the detection limit of a certain element, we assigned a random number chosen from a discrete uniform distribution, from zero to the minimum detection limit in parts per million (ppm) or percent. We used a random number drawn from the discrete uniform distribution because the sample could have contained slightly less than the detection limit, or the sample could have contained significantly less than the detection limit. This ensures the overall shape of the element's distribution does not change and reduces bias from a large number of samples using the detection limit as their analysis value.

To address the data issue of missing elemental values from the sediment and rock datasets, we used multiple imputation using chained equations (MICE). MICE has been used in psychology and medicine (Azur and others, 2011; White and others, 2011) and is a statistical technique used to fill in missing data. In general, MICE uses the distribution of observed data to fill in values for the missing data based on the assumption that the absent values are missing at random. Azur and others (2011) provide a more detailed description of the steps used in MICE:

1. Fill all missing values with random values used as placeholders.
2. Set placeholder values to missing for the first element to fill in.
3. The observed values for the first element are regressed on the other variables in the dataset.
4. The missing values for the first element are filled in with predictions from the regression model. The filled-in and observed values are then used in the regression for other elements.
5. Steps 2–4 are repeated for each element in the dataset for a specific number of cycles.

We used MICE for both the sediment and rock datasets to independently determine the missing values after 100 cycles. Upon completion, both the NURE sediment and NGDBR rock datasets contained a complete geochemical suite. We validated this method using a subset of both datasets that have all analytical values for each element. We then systematically removed certain known values to determine if MICE could reproduce the original dataset. Using a Kolmogorov-Smirnov (K-S) test, the filled-in values are statistically the same as the actual values ($p > 0.1$) for each element (table 1) and provided select elements for visual inspection (fig. 4). In figure 4, we cross plotted the filled in values with the actual values, and report the mean absolute error and absolute error of the filled-in values. With a full geochemical suite for sediment and rock samples, we next evaluated the spatial distribution of elements across subwatersheds.

Table 1. Results of imputation for each element in the NURE dataset.

Element	Crustal Abundance (ppm)	Dataset Element Mean (ppm)	Imputed Element Mean (ppm)	MAE (ppm)	K-S Test Statistic	K-S p-value
Ag	0.07	1.60	1.50	1.46	0.03	1
Al	82,300	5.40	5.40	0.99	0.07	0.72
As	1.80	4.30	4.70	2.39	0.31	0.35
Au	0.002	0.003	0.001	0	0.01	0.99
B	10	21	21	10.15	0.19	0.26
Ba	425	677	707	200	0.08	0.54
Be	2.80	1.70	1.60	0.44	0.02	1
Bi	0.17	2.50	2.50	2.40	0.06	1
Ca	41,500	25,000	24,000	16,100	0.08	0.45
Cd	0.20	2	1.80	1.70	0.04	1
Ce	60	64	65	22.89	0.06	0.88
Co	25	8	8.60	2.91	0.07	0.63
Cr	100	70	67	29.89	0.06	0.88
Cs	3	2.90	2.70	1.59	0.07	0.81
Cu	55	19	19	9.03	0.06	0.79
Dy	3	4	4	1.12	0.03	1

Table 1 continued.

Element	Crustal Abundance (ppm)	Dataset Element Mean (ppm)	Imputed Element Mean (ppm)	MAE (ppm)	K-S Test Statistic	K-S p-value
Eu	1.20	1.20	1.20	0.30	0.07	0.81
Fe	56,300	20,000	21,000	6,500	0.07	0.63
Hf	3	9.60	10	6.14	0.08	0.45
K	20,900	14,000	14,000	3,300	0.10	0.31
La	30	38	38	12.43	0.07	0.62
Li	20	28	28	10.54	0.06	0.88
Lu	0.50	0.30	0.30	0.15	0.07	0.89
Mg	23,300	16,000	16,000	6,400	0.04	0.99
Mn	950	427	448	176.86	0.05	0.97
Mo	1.50	1.70	1.50	1.56	0.07	1
Na	23,600	9,000	9,000	3,600	0.10	0.25
Nb	20	10	10	7.54	0.10	0.25
Ni	75	16	17	11.04	0.05	0.97
P	1,050	698	691	265	0.15	0.53
Pb	12.50	12	12	7.72	0.07	0.62
Rb	90	39	43	21.07	0.15	0.10
Sb	0.20	0.50	0.50	0.66	0.06	0.95
Sc	22	7.40	7.40	2.19	0.07	0.72
Se	0.05	1.60	1.80	1.13	0.12	1
Sm	6	5.60	5.10	2.16	0.07	0.81
Sn	2	4.10	8.40	7.41	0.20	0.11
Sr	375	261	271	127.69	0.08	0.54
Ta	2	0.19	0.08	0.18	0.05	0.98
Tb	0.90	0.13	0.14	0.18	0	1
Th	9.60	12.00	13	6.62	0.05	0.97
Ti	5,700	3,140	3,077	1,027	0.06	0.88
U	2.70	3.40	3.30	3.64	0.09	0.47
V	135	78	74	26.08	0.10	0.25
W	1.50	7.60	7.40	5.11	0.06	0.95
Y	33	14	14	4.42	0.05	1
Yb	3	3.10	2.80	1.22	0.07	0.89
Zn	70	60	57	33.79	0.05	0.93

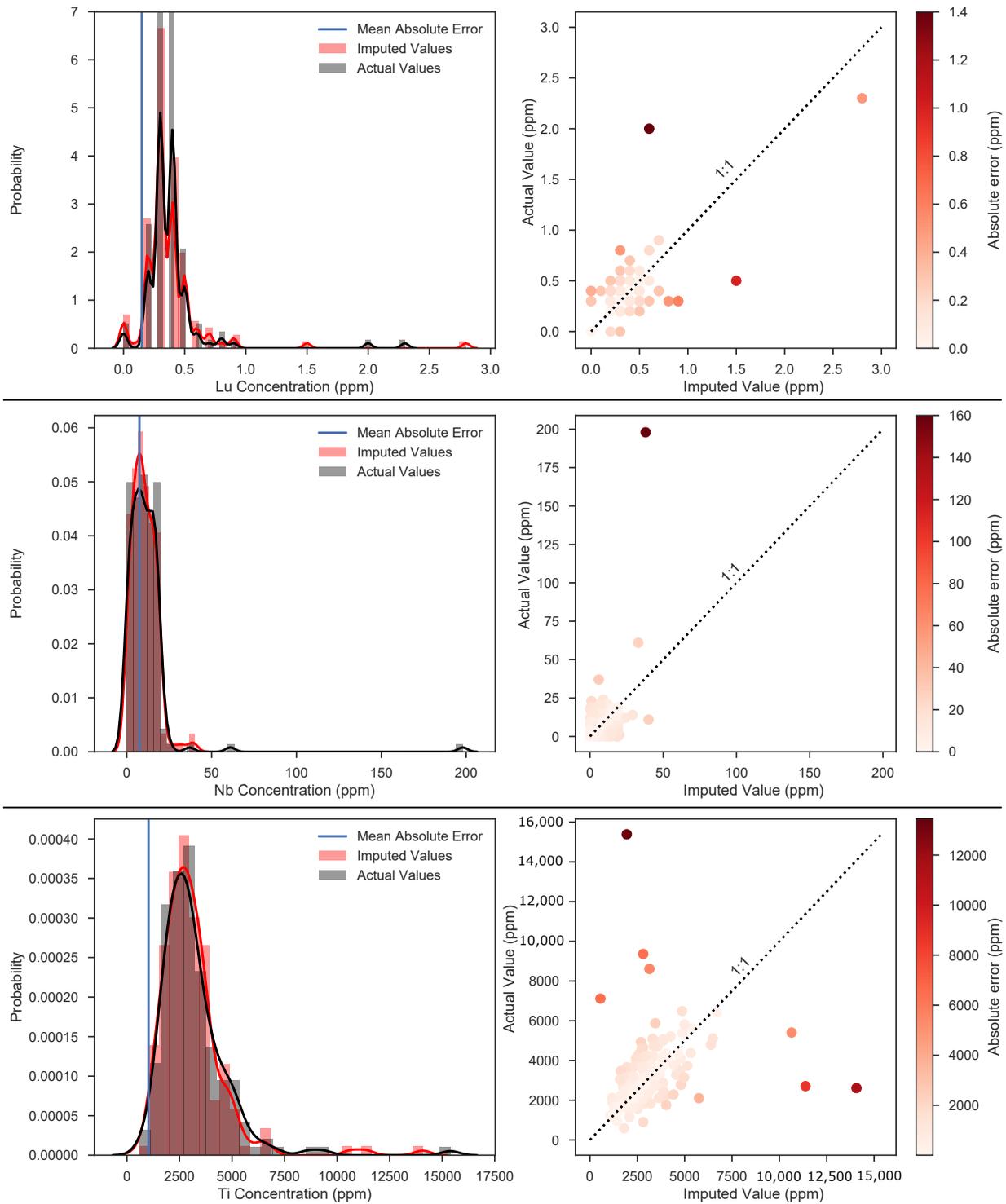


Figure 4. Comparison of imputed and actual values for lutetium (Lu), niobium (Nb), and titanium (Ti). The left column compares the distribution of the imputed values and actual values. The right column is a crossplot of actual and imputed values colored by absolute error. Dashed line has a 1:1 slope and showing perfect agreement.

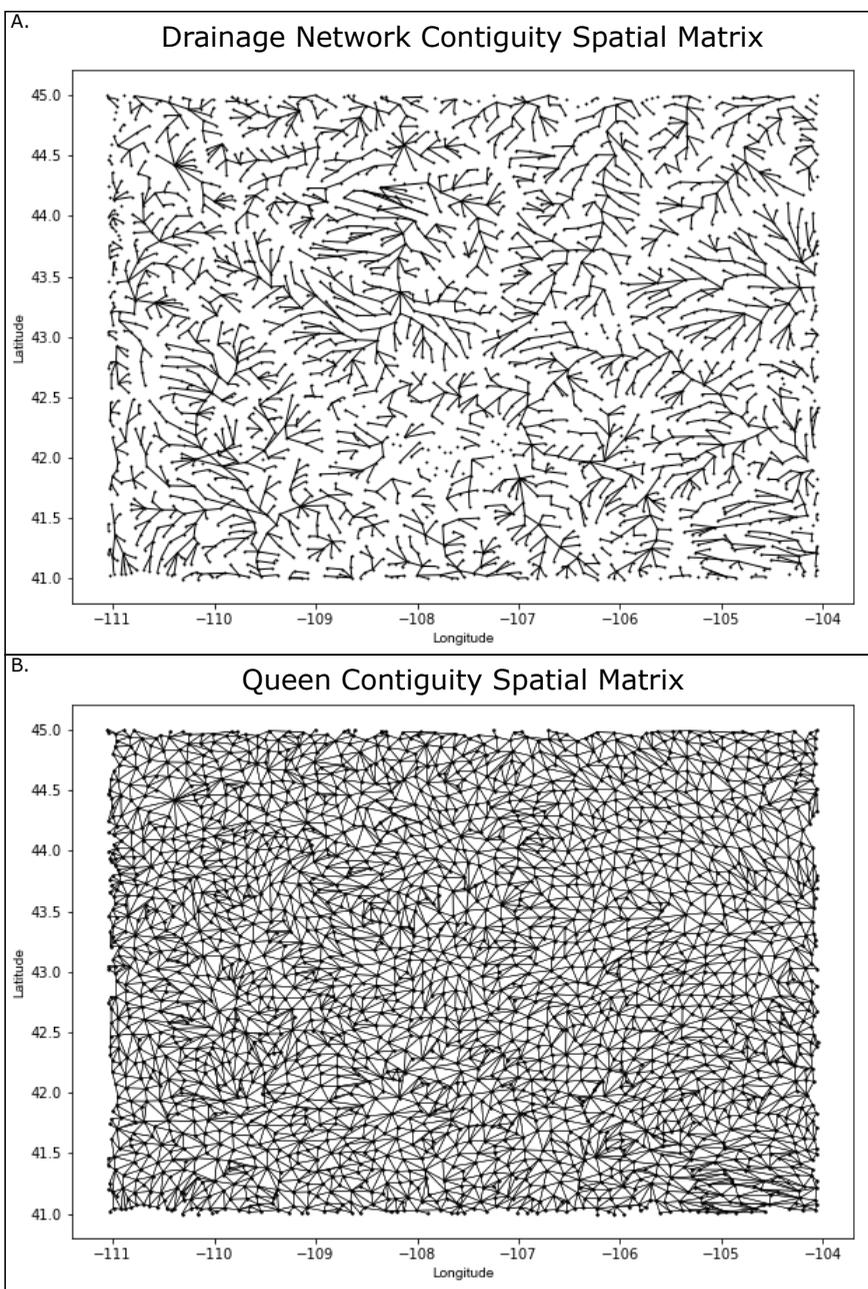
Spatial Associations

For this next step, we integrated the sample locations with the subwatershed dataset. The rock and sediment data were binned by each subwatershed. The maximum values for each element were then assigned to the subwatershed basin for both sediment and rock samples. We chose to bin the point data to polygons since we were covering the entirety of the state. This is different from methods used by Yager and others (2013), who analyzed the NURE sediment samples as point data. Although we used a similar spatial-hotspot analysis, we used a contiguity-weight spatial matrix, and binning the samples by subwatershed mitigates oversampling bias in prospective regions. A contiguity-weight spatial matrix is a specific type of spatial weights matrix used to represent the spatial structure of a dataset. Specifically, it is defined as a connection between two adjacent polygons.

For the NURE sediment dataset, we used a drainage network contiguity weight scheme: subwatersheds up and downstream of one another are considered neighbors, while subwatersheds on opposite sides of drainage divides are not. We used this neighborhood scheme under the assumption that sediment is only transported downstream and that adjacent subwatersheds across a drainage divide are not connected in the sense of sediment transport (fig. 5a). Once we constructed spatial weights matrices for both the sediment and rock datasets, we could further investigate differences or similarities between connected subwatersheds and look for patterns in the spatial data.

For the NGDBR dataset, we used a queen contiguity-weight scheme where if two subwatersheds share any boundaries (including single points), they are considered neighbors. We used this scheme for the rock-sample dataset because we assume that if altered rock is present in one subwatershed, then it is potentially also present in an adjacent subwatershed (fig. 5b).

Figure 5. Maps of contiguity-spatial networks used in hotspot analysis. Each subwatershed is represented by a point, and connectivity is represented by a line connecting the points. (A) Drainage spatial contiguity network graph used in the NURE sediment-hotspot analysis where subwatersheds are connected by geomorphic drainage patterns. (B) Queen spatial-contiguity network graph used in the NGDBR rock hotspot analysis where all touching subwatersheds are connected.



Getis-Ord G_i^* Hotspot Analysis

To analyze the spatial distribution of each element in the subwatershed across the state, we employ a spatial statistic used to document spatial clustering or anti-clustering of values in two-dimensions. This statistic is called the Getis-Ord G_i^* , and it documents areas of elevated or lowered values relative to a global mean. The Getis-Ord G_i^* statistic was developed by Getis and Ord (Getis and Ord, 1992; Ord and Getis, 1995) as a metric to document spatial clustering of a feature. Formally, it is defined as:

$$G_i^* = \frac{\sum_{j=1}^n w_{i,j} x_j - \bar{X} \sum_{j=1}^n w_{i,j}}{S \sqrt{\frac{n \sum_{j=1}^n w_{i,j}^2 - \left(\sum_{j=1}^n w_{i,j} \right)^2}{n-1}}}$$

where x_j is the attribute value for feature j , $w_{i,j}$ is the spatial weight between feature i and j , n is the total number of features,

$$\bar{X} = \frac{\sum_{j=1}^n x_j}{n}$$

and

$$S = \sqrt{\frac{\sum_{j=1}^n x_j^2}{n} - (\bar{X})^2}.$$

In this case, Getis-Ord G_i^* compares the value of an element in a subwatershed to the value of the same element in neighboring subwatersheds. If the subwatershed has a high value for that element and the neighboring features also have similarly high elemental values, the calculated G_i^* value will be high. When the local sum is greater than the expected local sum, it is statistically significant. We calculated the G_i^* statistic, considered a hotspot analysis, for both the sediment- and rock-sample datasets independent of one another.

G_i^* Values

The result from the hotspot analysis (G_i^*) is a value that spans all real numbers. G_i^* values can be thought of as a distance from the mean of the distribution. A G_i^* value of +2.0 is the same as a value located two standard deviations from the mean (2σ). Another way to think of a G_i^* value is as the cumulative percent of the distribution. Values less than -2.0 and greater than +2.0 document the 2.3 percentile and 97.7 percentile, respectively. Accordingly, G_i^* values less than -2.0 and greater than +2.0 are statistically significant ($p < 0.05$). Negative G_i^* values document statistical cold spots, or areas of low elemental abundance, while positive G_i^* values document statistical hotspots, or areas of high elemental abundance. While it is useful to know where there are hot and cold spots for each element, to be most effective we need to use the sediment and rock data in concert together when available. Because G_i^* values document the relative abundance of each element for both datasets, we can classify subwatersheds by the sediment and rock G_i^* values.

Classification

Each subwatershed with both a rock and sediment sample are classified by G_i^* values into four distinct classes (fig. 6). The first class consists of rock samples with G_i^* values above +2.0 and sediment samples ranging from -2.0 to +2.0. We call this class “known areas” because of the a posteriori knowledge that rock samples were deliberately collected from areas that are statistically hot. We are less concerned with sediment in these cases because we have ground-truth evidence of a hotspot from the rock samples.

The second class consists of sediment samples with G_i^* values above +2.0 and rock samples less than +2.0. We call this class “missed areas” because while the rock samples do not indicate a hotspot, the sediment samples do. This means that barring any major drainage reorganization since the Holocene, there is opportunity to find the outcrops that are the source to the above-average sediments.

The third class consists of sediment and rock samples with G_i^* values less than +2.0 but greater than -2.0. We call this class “background” with the interpretation that all neutral or average samples for both rock and sediment fall into this category.

The fourth class consists of sediment and rock samples with G_i^* values less than -2.0. We call this class “below background” because the samples in this class indicate cold spots for both sediment and rock samples. This means that the element of interest is depleted in these areas.

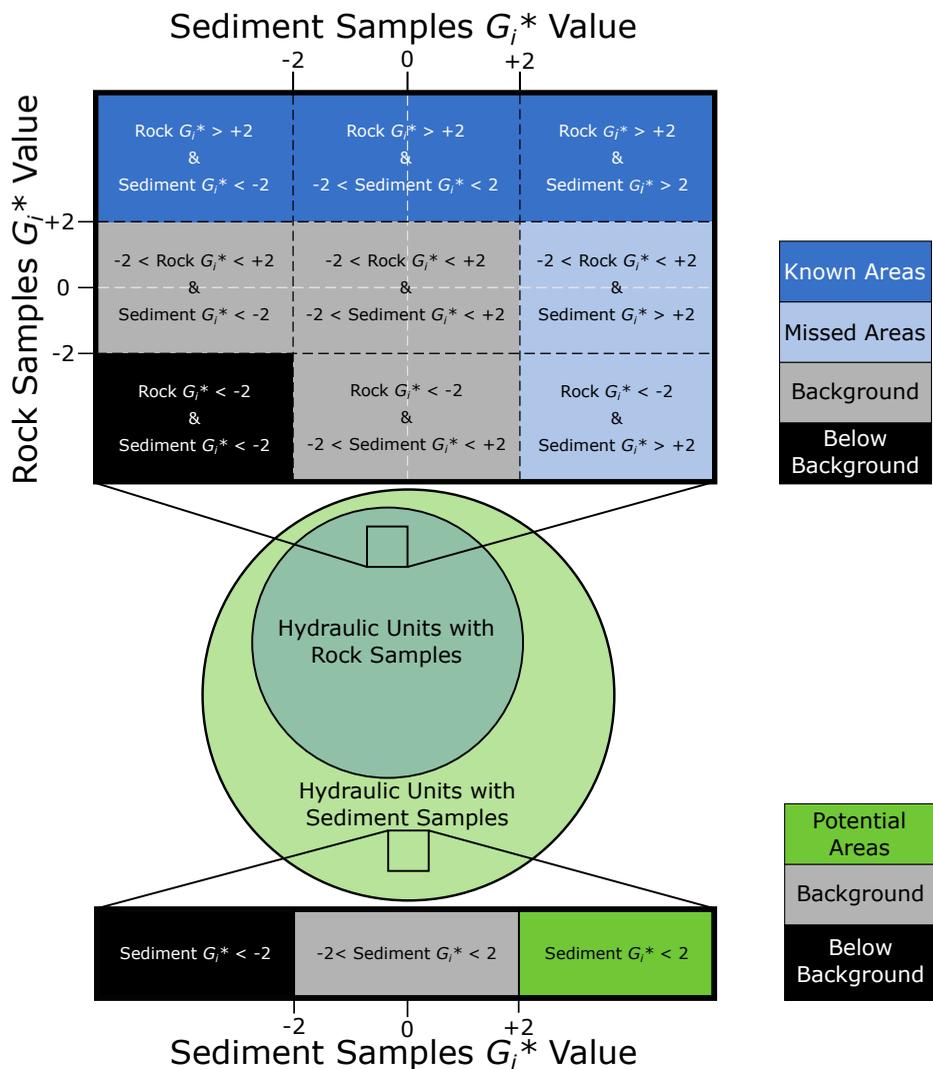


Figure 6. Visual representation of the classification scheme used to symbolize known, missed, potential, average, and below background areas. In the center of the figure is a Venn diagram documenting subwatersheds with sediment and rock samples. As evident from the Venn diagram, there are no subwatersheds with only rock samples and no sediment samples. If a subwatershed contains a sediment sample and rock sample, it is plotted in the two-dimensional space at the top of the figure. The G_i^* values for both the sediment and rock sample in the subwatershed is then used to classify the subwatershed. If a subwatershed contains only a sediment sample, it is plotted in the one-dimensional space at the bottom of the figure, and the subwatershed is classified by its sediment sample G_i^* values.

In cases where there are only sediment samples and no rock samples within a subwatershed, we used a three-class classification scheme. Subwatersheds with sediment sample G_i^* values of less than -2.0 are classified as “below background” in that the G_i^* values document a cold spot in the sediment samples (fig. 6). Subwatersheds with sediment samples between -2.0 and +2.0 are classified as “background” as they are similar to the “background” class listed above. These subwatersheds are neutral and do not document a hot or cold spot. Finally, subwatersheds with sediment sample G_i^* values of +2.0 and greater are classified as “potential areas” because they document statistically hot areas where there are no rock samples to constrain the source of the sediment. There are no subwatersheds that contain only rock samples and no sediment samples. The most important classes for the purpose of this report are the “known,” “missed,” and “potential” areas as they highlight areas of interest for future exploration.

Classification Validation Workflow

To test the validity of the classification model and data completion methods, we focus on titanium and rare earth elements (REEs). We chose titanium and REEs because of the published rock geochemical data for REEs and titanium in Wyoming by Sutherland and others (2013) and Sutherland and Cola (2015, 2016).

The workflow for titanium is as follows:

1. Fill in all missing element values for the sediment samples using MICE.
2. Bin the sediment-sample values to hydraulic unit subwatersheds and select the highest sample value for titanium to represent each hydraulic unit.
3. Build a dendritic spatial contiguity matrix for the subwatersheds to analyze the sediment samples.
4. Calculate G_i^* for each subwatershed using the binned titanium values.
5. Fill in all missing element values for the rock samples using MICE.
6. Bin the rock-sample values to hydraulic unit subwatersheds and select the highest sample value for titanium to represent each hydraulic unit.
7. Build a queen spatial contiguity matrix for the subwatersheds to analyze the rock samples.
8. Calculate G_i^* for each subwatershed using the binned titanium values.
9. Classify each subwatershed according to the sediment- and rock-sample G_i^* values.
10. Classify each subwatershed without rock samples according to the sediment-sample G_i^* values.

We then visually inspect the results of the classification and compare descriptive statistics for each class across the entire state.

The workflow for the REEs is similar to the workflow for titanium, with the exception that after we calculate the G_i^* values for each individual element at the subwatershed level (step 8), we sum the G_i^* values for all REEs within each subwatershed. Because we sum the G_i^* values, any subwatershed with REEs that has a value above or below our ± 2.0 threshold will be classified in the same way as individual elements.

RESULTS

Titanium

Areas with sediment sample G_i^* values greater than +2.0 are randomly distributed across the state with some clusters in the Greater Green River Basin, the central Laramie Mountains, and a few areas near the Black Hills Uplift (fig. 7a). Areas in Wyoming that have rock sample G_i^* values greater than +2.0 for titanium include the Black Hills Uplift in the northeast, the Absaroka Range in the northwest, the west-central Gros Ventre Range, the northwest part of the west-central Wind River Range, and the Medicine Bow Mountains and northern Laramie Mountains in the southeast (fig. 7b). In total, the potential areas cover 19,200 km² (7.5 percent of the state), the missed areas cover 12,186 km² (4.8 percent of the state), and the known areas cover 14,839 km² (5.8 percent of the state). All combined, the known, missed, and potential areas constitute 46,266 km² or 18.1 percent of the state (fig. 7c).

Rare Earth Elements

Regions with sediment sample G_i^* values greater than +2.0 are much more randomly distributed than the rock-sample hotspots, but we do note continuous hot areas both north and south of the Wind River Range, in the Absaroka Range, and in the central Laramie Mountains (fig. 8a). Areas in Wyoming that have rock sample G_i^* values greater than +2.0 for all REEs include the Black Hills Uplift in the northeast (specifically the area surrounding the Bear Lodge Mountains), the Absaroka Range in the northwest, the centrally located Granite Mountains, the west-central Gros Ventre and Wind River ranges, and the Medicine Bow and Sierra Madre mountains in the southeast (fig. 8b). The highest number of missed areas, not surprisingly, tend to surround the known areas (fig. 8c). In total, the potential areas cover 19,990 km² (7.8 percent of the state), the missed areas cover 12,365 km² (4.8 percent of the state), and the known areas cover 31,385 km² (12.3 percent of the state). All combined, the known, missed, and potential areas constitute 63,740 km² or 24.9 percent of the state. This G_i^* values classification scheme correctly identifies the Bear Lodge REE project in the northeast corner of the state as a known deposit, with some missed areas to the northwest of the deposit (fig. 8c).

Validation

To further validate the G_i^* values classification method, we used the ground-truth rock-chip samples from Sutherland and others (2013) and Sutherland and Cola (2015, 2016) to provide a baseline for prediction accuracy. We used all 418 samples to validate our classification scheme for subwatersheds. Our validation compares the predicted class of each subwatershed to a binary classification of the samples collected and analyzed by Sutherland and others (2013). Known, missed, and potential areas are classified as positive predictions, and their count is summed. Background and below background areas are classified as negative predictions, and their count is summed. Using this binary classification scheme, we then calculate the numbers of true positives (correctly predicted known, missed, and potential areas), true negatives (correctly predicted background and below background areas), false positives (incorrectly predicted known, missed, and potential areas), and false negatives (incorrectly predicted background and below background areas). From the counts for the predicted and actual classes we then calculate accuracy, sensitivity, and specificity of the model for 1–5 times crustal abundance of elements (Taylor and McLennan, 1995; table 2).

Accuracy is defined as the sum of true positives and true negatives divided by the sum of the total population, or simply the number of correct predictions divided by the total number of predictions. Sensitivity is the sum of true positives divided by the sum of actual positives, while specificity is the sum of true negatives divided by the sum of actual negatives. A model with high sensitivity and low specificity has few false negatives and many false positives, meaning the model would predict missed or potential areas where they are actually background or below background areas. A model with low sensitivity and high specificity has many false negatives and few false positives, and would predict background or below background areas when they are actually missed or potential areas. We prefer a model with high sensitivity and low specificity as it is more important to overestimate the mineralization potential of an area and verify the modeled predictions with a sampling program than to underestimate the potential for mineralization and never disprove the model with a sampling program.

Figure 7. Map showing results from hotspot analysis and classification for titanium. The maps document the following: (A) G_i^* values for sediment samples (NURE dataset), (B) G_i^* values for rock-chip samples (NGDDBR dataset), and (C) classification results for both rock-chip and sediment samples. Dashed rectangle to the southeast is fig. 9a, and dashed rectangle to the southwest is fig. 10a.

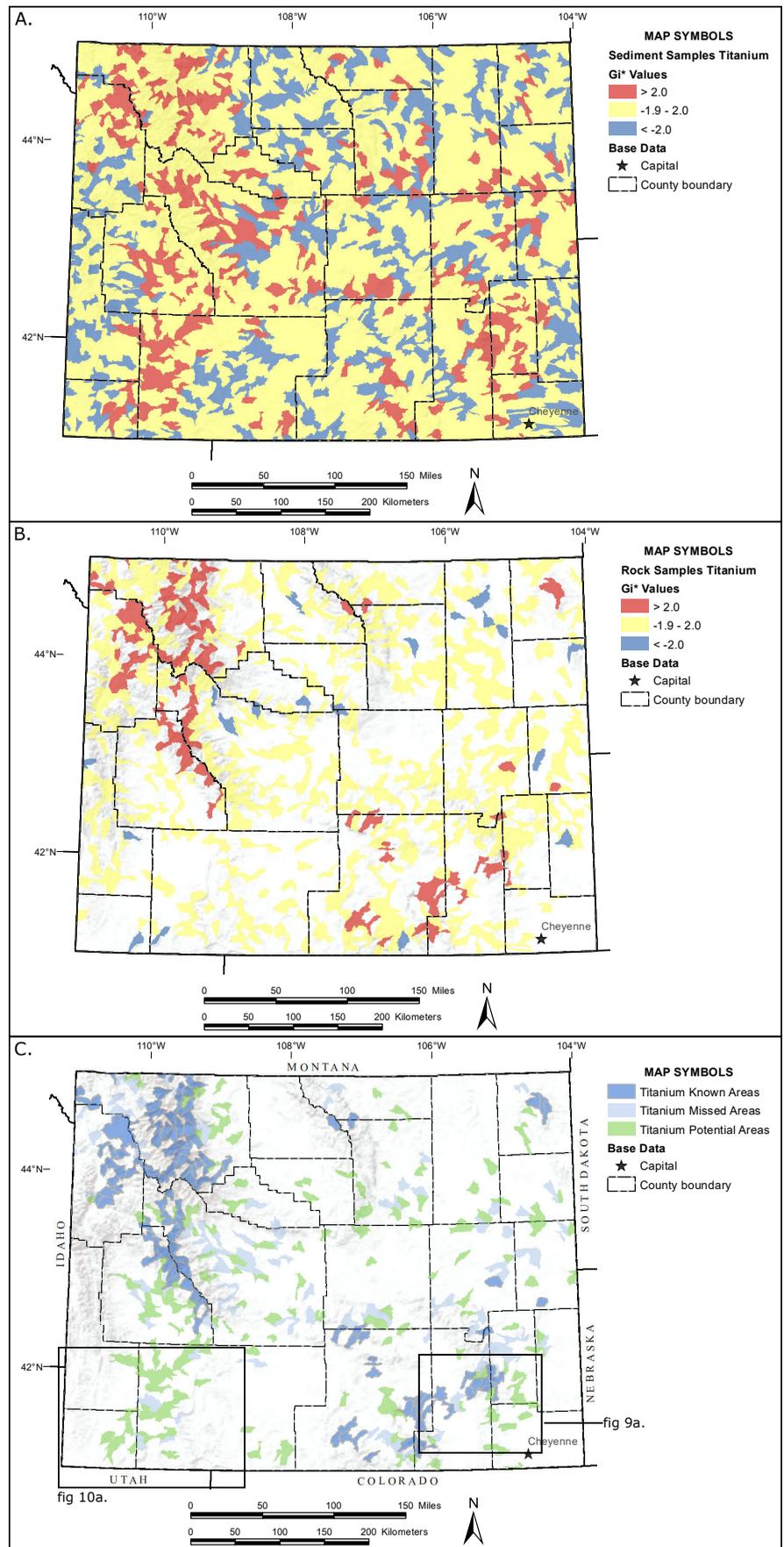


Figure 8. Map showing results from hotspot analysis and classification for all rare earth elements. The maps document the following: (A) G_i^* values for sediment samples (NURE dataset), (B) G_i^* values for rock-chip samples (NGDBR dataset), and (C) classification results for both rock-chip and sediment samples. Dashed rectangle to the northeast is fig. 9b, and dashed rectangle to the southwest is fig. 10b.

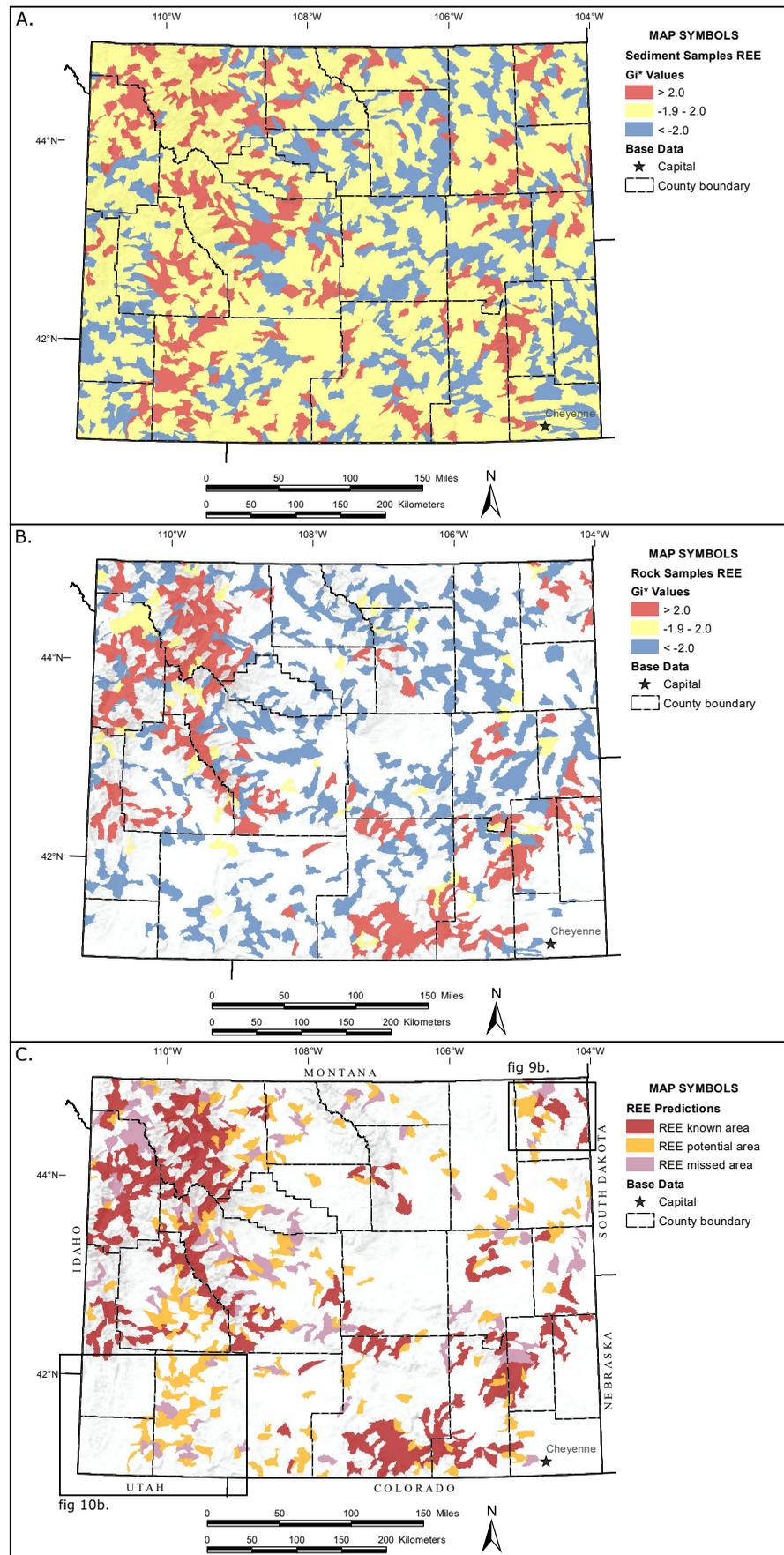


Table 2. Accuracy, sensitivity, and specificity of the classification tested on the data from Sutherland and others (2013) and Sutherland and Cola (2015, 2016).

Concentration of Elements Relative to Crustal Abundance	Accuracy		Sensitivity		Specificity	
	Titanium	REEs	Titanium	REEs	Titanium	REEs
1x	74.4%	53.1%	0.18	0.36	0.80	0.68
2x	77.3%	57.4%	0.17	0.31	0.80	0.65
3x	76.6%	60.3%	0.11	0.31	0.80	0.65
4x	79.2%	60.3%	0.17	0.26	0.80	0.65
5x	70.1%	60.0%	0.21	0.33	0.80	0.66

From the rock and sediment samples for titanium we note that the accuracy of the classification increases from 74 percent to 79 percent as the concentration of elements relative to crustal abundance increases, up to four times crustal abundance, before dropping to 70 percent accuracy at five times crustal abundance (table 2). The sensitivity of the classification scheme remains constant with little variation relative to crustal abundance (± 0.1) as element concentration increases. The specificity of the classifier remains constant with no variance across different crustal abundances, implying the model classifies background areas as known, missed, or potential more often than it classifies known, missed, or potential areas as background. We interpret the model behavior of classifying a subwatershed with average concentrations of titanium as being a known or missed area as more desirable than classifying an area with elevated levels of titanium as average.

From the rock and sediment samples for REEs we note a similar trend as with titanium. The accuracy of the classifier increases with the concentration of elements relative to crustal abundance up to three times crustal abundance, then remains constant to five times crustal abundance (table 2). The sensitivity of the classifier remains constant with some variation (± 0.1) as element concentration increases. The specificity of the classifier remains constant with little variance across different crustal abundances (± 0.03). The low-sensitivity values and high-specificity values for all levels of crustal abundance mean the model is more likely to have more false negatives and fewer false positives. As with titanium, we suggest the model behavior of classifying a subwatershed with average concentrations of REEs as being a known or missed area is preferable to classifying an area with elevated levels of REEs as average.

Case Studies

In this section, we cover two areas of interest for titanium and rare earth elements. We used the classification scheme to identify areas of interest for titanium and REEs, and relate results from previous geological studies to the model predictions.

Titanium

We used the northern Laramie Mountains as a case study area for titanium because the area contains previously documented titaniferous iron deposits (Sutherland and Cola, 2015). In the northern Laramie Mountains, the model predicts known and missed areas covering and surrounding the titaniferous deposits. Additionally, the subwatersheds with only sediment samples predict potential areas of titanium that remain unsampled for rock samples (fig. 9a). The known areas surrounded by missed areas and potential areas is what we would expect to see in areas where titanium erodes from a mineralized area and is transported downstream. The potential areas in the northern Laramie Mountains are of particular interest to constrain additional rock sources of titanium in the drainages.

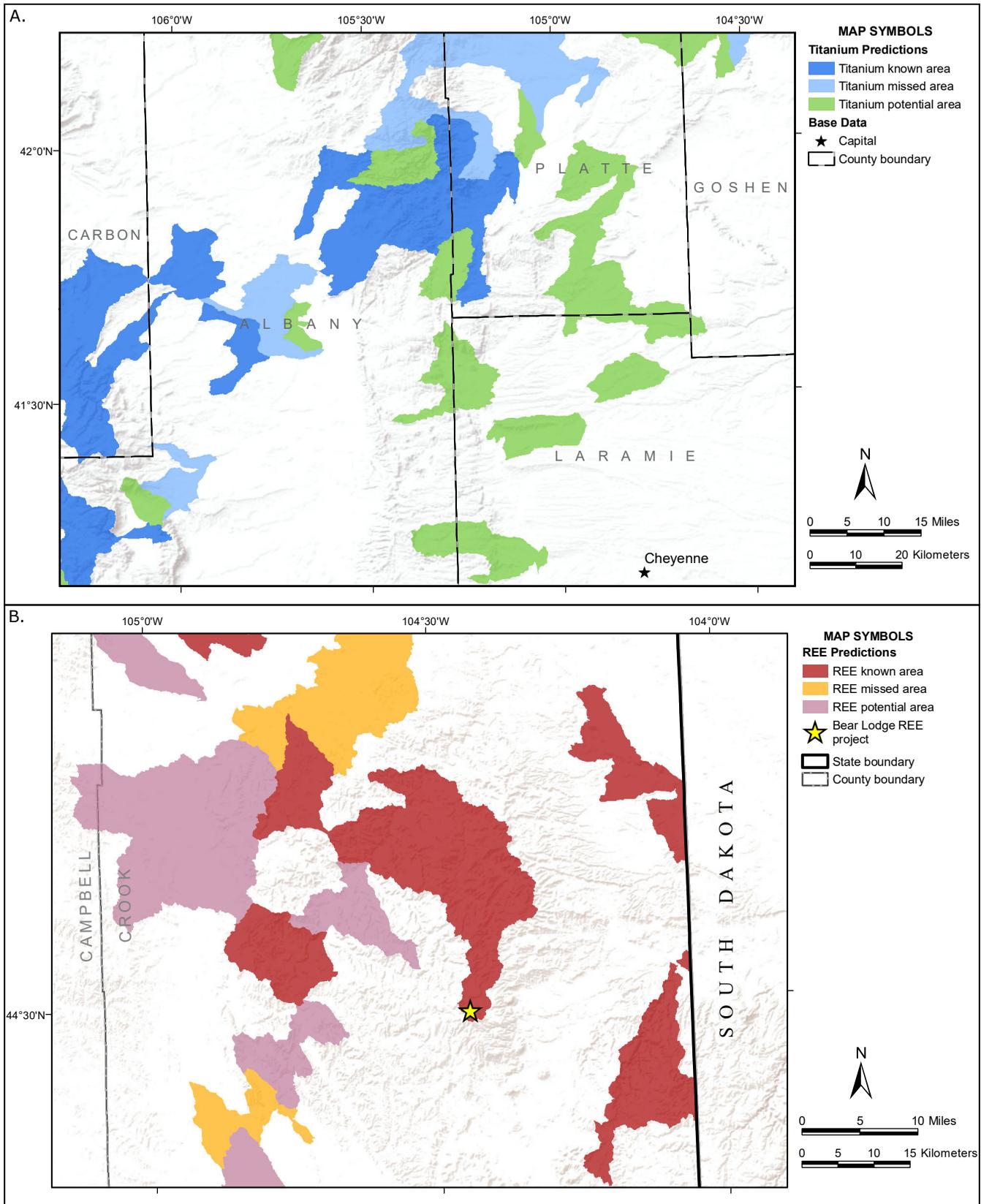


Figure 9. Map of results from case studies, including (A) known, missed, and potential areas of interest for titanium in the Laramie Mountains and (B) known, missed, and potential areas of interest for REEs in the Bear Lodge REE district in the Black Hills Uplift.

Rare Earth Elements

The Bear Lodge REE project in the Bear Lodge Mountains along the western limb of the Black Hills Uplift is an area with known REE development potential (Sutherland and others, 2013; Sutherland and Cola, 2016). We used it as a case study to examine the validity of our classification scheme. The subwatersheds that contain sampled portions of the deposit are classified as known areas based on sediment and rock sample G_i^* values, implying the model can correctly identify areas with known deposits. To the west of the known mineralized area, the sediment samples are classified as potential areas for REEs. This is most likely because the REEs weathering out of the Bear Lodge district are transported to the west by rivers. However, because the center of the deposit was not sampled and not included in our dataset, it is labeled as background. In this case, the sediment samples provide a vector of areas to focus on, and they direct us to the primary deposit in the Bear Lodge Mountains (fig. 9b).

FUTURE SAMPLING

At the statewide scale, one area of potential exploration interest for both REEs and titanium indicated from the sediment data is the southwest corner of the Greater Green River Basin (fig. 10). There are several subwatersheds classified as potential areas, and no associated rock-sample data exist for the area to validate the sediment-sample data. Additional rock sampling of the subwatersheds in this area would be useful to verify the anomalous sediment samples and to potentially pinpoint the source of the anomalies (fig. 10). Hausel (1998) and Sutherland and Cola (2016) discuss the potential of breccia pipes in this part of the state as hosts of REEs, lending further credence to the model predictions of missed and potential areas. More detailed geologic field work is needed to validate the model predictions in these areas for both REEs and titanium. However, the model works well as a regional exploration tool and is easily applied to all elements in the dataset.

SUMMARY AND CONCLUSIONS

In this report we present a novel use of multiple imputation using chained equations (MICE) to fill in missing values in a geochemical dataset. We validated the method by creating a test dataset for each element and calculating the mean absolute error. A K-S test validated the similarity of the dataset before and after filling in missing values. Once we filled in the missing values, we spatially binned the highest values for each element to hydraulic units. After spatial binning, we calculated the Getis-Ord G_i^* statistic for both sediment and rock sample datasets. This statistic documented hot and cold spots for each element across the state. The resulting G_i^* values for sediment and rock datasets were then used to classify subwatersheds as known, missed, potential, background, or below background. The classification model, validated against 418 samples for both rare earth elements and titanium, documents that this G_i^* values classification scheme tends to overestimate the number of missed and potential areas. We suggest that searching a wider area of possibilities is preferable to limiting the search area and possibly missing a mineralized area.

The primary use of this model is to highlight areas in Wyoming for focused sampling programs for specific minerals. It also is useful for regional-scale exploration of critical and strategic minerals and resources. In addition, this model will aid in selecting areas for further geologic investigation through mapping. Although we present two case studies for rare earth elements and titanium that confirm the modeled predictions, future work on critical and strategic elements and minerals across the state is needed to validate the model for different elements. The benefit of using hotspot analysis with classification is that this method is applicable at different scales. This workflow and model are easily applied at both the local, regional, and national scale. Classification and hotspot results for all elements are accessible through the WSGS [Mines and Minerals Map](http://wsgs.maps.arcgis.com/apps/webappviewer/index.html?id=af948a51f4954a81adeae8935440cd28) (http://wsgs.maps.arcgis.com/apps/webappviewer/index.html?id=af948a51f4954a81adeae8935440cd28) and through the WSGS [publication search page](https://www.wsgs.wyo.gov/pubs-maps/publication-search) (https://www.wsgs.wyo.gov/pubs-maps/publication-search).

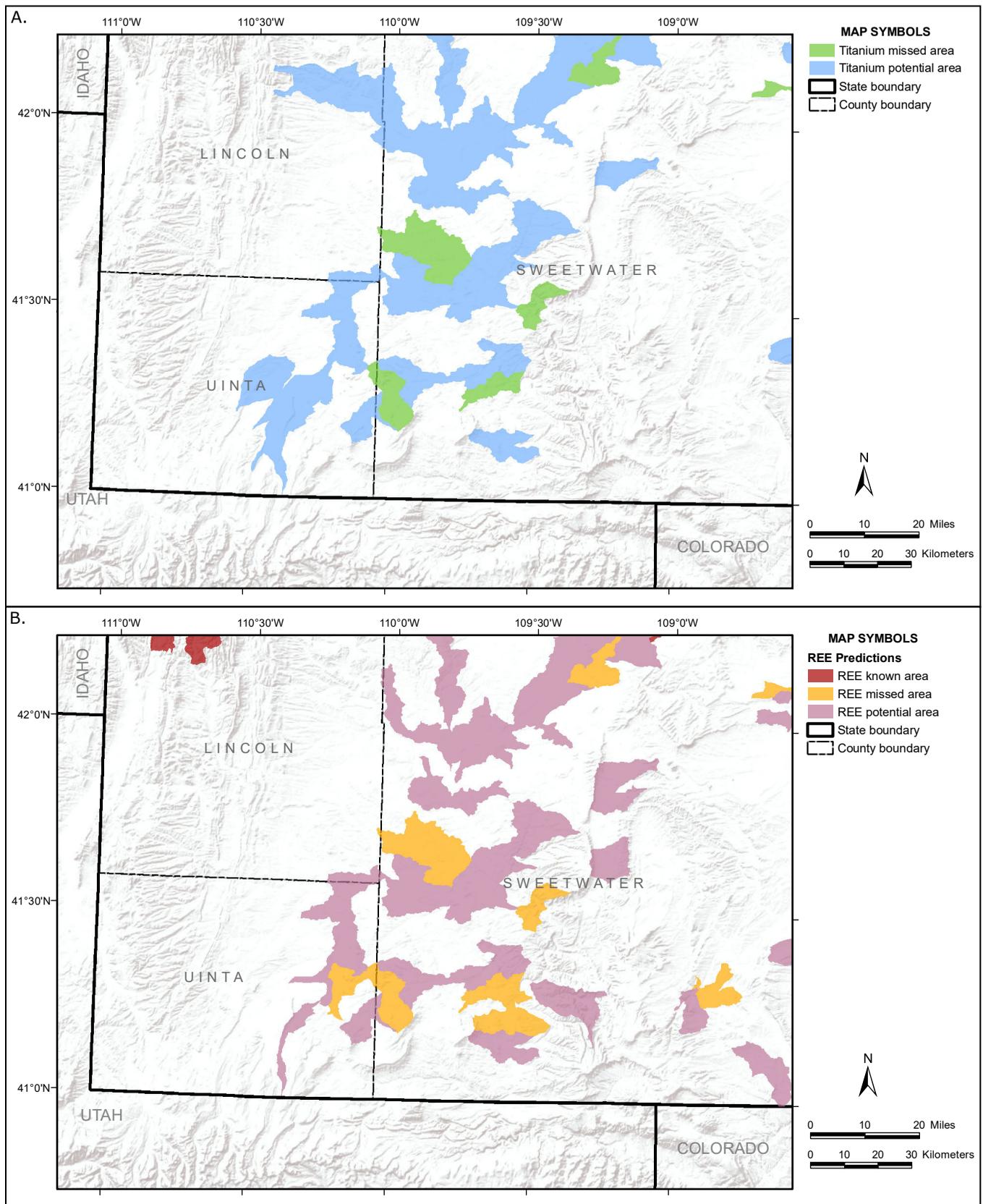


Figure 10. Map of case-study results from the Greater Green River Basin, including (A) missed and potential areas of interest for titanium and (B) missed and potential areas of interest for REEs. The titanium and REE missed and potential areas have spatial overlap and may be sourced from a yet-unknown mineralized area.

REFERENCES

- Azur, M.J., Stuart, E.A., Frangakis, Constantine, and Leaf, P.J., 2011, Multiple imputation by chained equations—What is it and how does it work?: *International Journal of Methods in Psychiatric Research*, v. 20, no. 1, p. 40–49.
- Getis, Arthur, and Ord, J.K., 1992, The analysis of spatial association by use of distance statistics: *Geographical Analysis*, v. 24, no. 3, p. 189–206.
- Hausel, W.D., 1998, Diamonds and mantle source rocks in the Wyoming craton with a discussion of other U.S. occurrences: *Wyoming State Geological Survey Report of Investigations 53*, 93 p., 3 pls.
- Ord, J.K., and Getis, Arthur, 1995, Local spatial autocorrelation statistics—Distributional issues and an application: *Geographical Analysis*, v. 27, no. 4, p. 286–306.
- Puchlik, Kenneth, 1977, Collection and preparation of wet and dry stream-sediment samples [In arid and semi-arid regions for National Uranium Resources Evaluation Program], United States, accessed March 2018, at <https://www.osti.gov/servlets/purl/7301137>.
- Sharp Jr., R.R., and Aamodt, P.L., 1978, Field procedures for the uranium hydrogeochemical and stream sediment reconnaissance as used by the Los Alamos Scientific Laboratory: Los Alamos, New Mexico, manual LA-7054-M, GJBX-68, 64 p.
- Smith, S.M., 1997, National Geochemical Database reformatted data from the National Uranium Resource Evaluation (NURE) hydrogeochemical and stream sediment reconnaissance (HSSR) program: U.S. Geological Survey, U.S. Department of the Interior, p. 97–492.
- Sutherland, W.M., and Cola, E.C., 2015, Iron resources in Wyoming: *Wyoming State Geological Survey Report of Investigations 67*, 91 p.
- Sutherland, W.M., and Cola, E.C., 2016, A comprehensive report on rare earth elements in Wyoming: *Wyoming State Geological Survey Report of Investigations 71*, 137 p.
- Sutherland, W.M., Gregory, R.W., Carnes, J.D., and Worman, B.N., 2013, Rare earth elements in Wyoming: *Wyoming State Geological Survey RI-65*, 82 p.
- Taylor, S.R., and McLennan, S.M., 1995, The geochemical evolution of the continental crust: *Reviews of Geophysics*, v. 33, no. 2, p. 241–265.
- U.S. Geological Survey, 2018, National Geochemical Database, accessed March 2018, at <https://mrddata.usgs.gov/ngdb/rock/map-us.html>.
- U.S. Geological Survey, and U.S. Department of Agriculture, Natural Resources Conservation Service, 2013, Federal Standards and Procedures for the National Watershed Boundary Dataset (WBD): *Techniques and Methods*, v. 11, p. 63.
- White, I.R., Royston, Patrick, and Wood, A.M., 2011, Multiple imputation using chained equations: issues and guidance for practice: *Statistics in Medicine*, v. 30, no. 4, p. 377–399.
- Yager, D.B., Hofstra, A.H., and Granitto, Matthew, 2012, Analyzing legacy U.S. Geological Survey geochemical databases using GIS—Applications for a national mineral resource assessment: *U.S. Geological Survey Techniques and Methods 11–C5*, 28 p.

